

UNIVERSITE DE POITIERS  
U.F.R. Sciences Fondamentales et Appliquées

Année 2004-2005

*Master Sciences et Technologies*  
*Mention Biologie Santé Agronomie*  
*Institut Universitaire Professionnalisé (I.U.P.)*  
*Génie Physiologique Informatique*

\*\*\*\*\*

1ère année de Master I.U.P.  
Rapport de stage

\*\*\*\*\*

# Statical Tools For Mixed Effect Modeling

## Programming Two New Functions in Splus

Effectué à :  
**Sanofi-Aventis Deutschland**  
Industriepark Höchst  
Frankfurt am Main  
du 01/06/2005 au 12/08/2005  
Sous la responsabilité de :  
**SPETH Heiner**

Présenté par :  
**MALARY Matthieu**

## Remerciements

Je remercie le **Docteur Jochen Mass**, responsable du département *Drug Metabolism and Pharmacokinetics* des sites de Francfort sur le Main et Paris, pour m'avoir permis de réaliser mon stage de 1<sup>re</sup> année de Master au sein du département et plus précisément dans le service *Pharmacocinétique et Pharmacodynamie des populations*.

Je remercie également les membres de ce service, **Docteur Willi Weber**, **Docteur Diether Rueppel** et **Heiner Speth** pour leur accueil et leur encadrement.

Mes remerciements vont aussi au professeur **Jean Michel Maixent** ainsi qu'à **Justine Lahaye** pour les contacts et renseignements apportés afin d'accomplir ce stage en Allemagne.

Enfin, je remercie mon coéquipier **Guillaume Gorson-Tanguy** pour sa collaboration tout au long de ces deux mois et demi.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Summary</b>	<b>6</b>
<b>3</b>	<b>Sanofi-Aventis</b>	<b>9</b>
3.1	Historique . . . . .	9
3.2	Médicaments et Produits . . . . .	10
3.3	Recherche et Développement . . . . .	11
3.4	Le département GMPK . . . . .	11
3.4.1	Organistaion . . . . .	12
3.4.2	Environnement humain . . . . .	12
3.4.3	Environnement technique . . . . .	12
<b>4</b>	<b>La pharmacocinétique/pharmacodynamie des populations</b>	<b>14</b>
4.1	La pharmacocinétique . . . . .	14
4.2	La pharmacodynamie . . . . .	14
4.3	Les paramètres utilisés . . . . .	15
4.4	Définition du modèle . . . . .	16
4.5	Les outils : . . . . .	17
4.5.1	NONMEM . . . . .	17
4.5.2	S-PLUS . . . . .	18
<b>5</b>	<b>Stage</b>	<b>19</b>
5.1	Présentation des trois méthodes existantes : . . . . .	19
5.1.1	Méthode Initial Thetas . . . . .	19
5.1.2	Méthode Bootstrap . . . . .	19
5.1.3	Méthode Jackknife . . . . .	20
5.2	Améliorations apportées : . . . . .	21
5.2.1	Les besoins : . . . . .	21
5.2.2	Modifications des applications : . . . . .	22
5.3	Nouvelles fonctions développées : . . . . .	25
5.3.1	Fonction Variation Onetheta . . . . .	26
5.3.2	Fonction Simplefit . . . . .	31
5.3.3	Génération des fichiers d'aide . . . . .	33
<b>6</b>	<b>Conclusion</b>	<b>35</b>
<b>7</b>	<b>Bibliographie</b>	<b>36</b>

<b>8 Annexes</b>	<b>37</b>
8.1 Control file : . . . . .	37
8.2 Data file : . . . . .	39

# 1 Introduction

Sanofi-Aventis a pour raison sociale, l'amélioration de la santé humaine par la prévention et le traitement des maladies. Société pharmaceutique numéro 1 en Europe, elle est spécialisée dans la recherche et le développement de molécule. Les effets thérapeutiques de ces molécules sont déterminés chez l'homme par des études cliniques. Le nombre d'individus inclus dans ces tests est limité pour des raisons éthiques, économiques et humaines. La posologie du médicament doit être déterminée afin d'obtenir le plus grand effet thérapeutique avec un effet toxique minimum pour le patient. Cette dose est établie sur une population par l'intermédiaire de simulation informatique à partir d'un échantillon. Cette méthode nécessite dans un premier temps, la création d'un modèle mathématique hypothétique simulant le devenir du médicament dans l'organisme. Ce modèle est corrélé à de nombreuses covariables. La définition de ces covariables et la simulation à partir des différents modèles sont réalisés au sein du service Pharmacocinétique/ Pharmacodynamie des populations dans lequel j'ai été intégré lors de mon stage de 1<sup>re</sup> année de Master.

L'informatique est un outil indispensable pour ces simulations. En effet, les données de départ et celles calculées sont nombreuses. De grandes quantités de fichier sont analysées grâce à la puissance actuelle des ordinateurs ce qui permet un rendu de résultat dans les plus courts délais et avec le minimum de biais. Pour cela, l'équipe utilise comme langage de programmation, le S-plus ainsi que le logiciel de regression mathématique NONMEM. Plusieurs fonctions basées sur des méthodes statistiques ont été développées l'an dernier par nos prédécesseurs (Justine Lahaye, Baptiste Moulinier et Matthieu Chosseler), tous trois anciens étudiants de l'I.U.P. "Génie Physiologique et Informatique".

Lors de ce stage, l'optimisation de ces méthodes ainsi que le développement de deux nouvelles nous ont été confiés à moi et à un autre stagiaire. Nous avons travaillé en commun afin de répondre aux besoins de l'équipe.

Après la présentation de l'entreprise, je rentrerai dans le détail de ce projet afin de vous présenter le travail accompli ainsi que les différentes technologies utilisées. Par la suite, je conclurai sur les apports de ce stage aussi bien sur le plan technique que humain.

## 2 Summary

### Sanofi-Aventis

#### An overview

Sanofi-Aventis is dedicated to treating and preventing diseases by discovering and developing innovative prescription drugs and human vaccines. This group is the first pharmaceutical company in Europe and the third worldwide. Sanofi-Aventis is present in more 100 countries with about 100000 employees. In 2004, the company generated 25 billion euros and invested 4 billion euros in research and development. The 7 major therapeutic areas of research are : cardiovascular, thrombosis, central nervous system, oncology, metabolic disorders, internal medicine and vaccines. The group was born from merge ,in April 2004, of two companies : Sanofi-Synthélabo et Aventis Pharma. At the origin, a public offer of purchase had launched by Sanofi-Aventis on its principal competitor Aventis. This one made a success and gave birth to the third world pharmaceutical group after Pfizer and GSK. Today, the company has 6 blockbusters : the anti-thrombosis agent Lovenox<sup>®</sup>/Clexane<sup>®</sup>, the antiplatelet agent Plavix<sup>®</sup>, the chemotherapy agent Taxotere<sup>®</sup>, the antihistamine agent Allegra<sup>®</sup>/Telfast<sup>®</sup>, the hypnotic agent Stilnox<sup>®</sup>/Ambien<sup>®</sup>/Myslee<sup>®</sup> and the anti-cancer drug Eloxatine<sup>®</sup>.

#### Research & Development

The research of the group has one of richest and innovating wallets of all world pharmaceutical industry with 128 molecules under development including 48 in advanced phases (phase II and III).

Present on 3 continents, 11 500 researchers work in more than 20 sites.

#### GMPK Department

I did training in the GMPK (*Global Drug Metabolism and Pharmacokinetics*) more precisely in the service interested in population Pharmacokinetic/Pharmacodynamic modeling. The team in Frankfurt is composed of three members : Willi Weber (physician, chemist, pharmacometrician), Dieter Ruppel (biophysicist, pharmacometrician) and Heiner Speth (computer scientist).

Population Pharmacokinetics refers to a method of pharmacokinetic data analysis that is especially suited to data from groups individuals, or population. The objective of population pharmacokinetic analyses is to estimate not only the typical (mean) pharmacokinetic parameters of a population but also

to obtain quantitative (and statistically correct) information on the extent and the sources of pharmacokinetic variability. This is useful for assessing the probability of drug concentrations that will be below or above the therapeutic range with a uniform dosing scheme, for setting up dose recommendations for subpopulations with altered pharmacokinetics, and as a basis for adaptive individual dosing in the case of drugs with a narrow therapeutic range. The team uses several statistical tools. The methods programmed in Splus launches the software of regression NONMEM on a model with initialized parameters. The final estimates returned by NONMEM are gathered in a report.

## **Training**

The prime objective of my training course was to optimize these methods with the needs of the team. Second was to develop two new functions correspondents with two new statistical approaches. This work was completed in collaboration with my fellow-member Guillaume Gorson-Tanguy.

## **Optimization**

The first part of our training course consisted in the discovery of the tools used and the langage of programming S-plus.

NONMEM is a mathematical regression software. This program uses much the resources of the computer. Moreover, one time launched, the treatment can last several hours. The problem is that NONMEM can “crash” and return null values. In this case, the generated report with null values is useless. We cut out the function in order to recover in this case the useful data and to avoid the generation of the report.

## **Two new functions**

In the second time, the team submitted to us the idea of two new functions.

The first function Variation.Onetheta is based on the function Initial.Thetas. This function varies the whole of the initial parameters in order to optimize the model. The goal of the new function is to study the variation of only one of the initial parameters. Moreover this function automatically starts again new calculations under certain conditions. Lastly, it allows the edition of the

results in the form of tables and graphs.

The second function Simplefit makes a statistical study on each individual of a sample. This one can be to compare to the other individuals and with the whole of the population.

These two functions were developed with Guillaume. We took care to document to the maximum the code and to create the various help files necessary to the user.



### 3 Sanofi-Aventis

Sanofi-Aventis a pour vocation de faire plus, mieux et plus vite dans le domaine de la santé. Elle s'appuie pour cela sur la découverte de molécules porteuses d'espoir ainsi que sur un accès aux médicaments pour le plus grand nombre.

Sanofi-Aventis est aujourd'hui l'entreprise pharmaceutique numéro 1 en Europe et numéro 3 au niveau mondial. Elle possède une part de marché mondial s'élevant à 5,3%. Le groupe est représenté dans le monde par près de 100000 collaborateurs dans plus de 100 pays sur les 5 continents. Le chiffre d'affaire généré en 2004 est de 25 milliards d'euros avec un budget alloué à la recherche et au développement de près de 4 milliards d'euros.

#### 3.1 Historique

Sanofi-Aventis est née de la fusion en avril 2004 de deux grands groupes Sanofi-Synthélabo et Aventis Pharma. A l'origine, Sanofi-Synthélabo avait lancé en janvier 2004 une offre publique d'achat sur son principal concurrent Aventis. Celle-ci a réussi et a donné naissance au troisième groupe pharmaceutique mondial après Pfizer et GSK. Sa présence internationale permet aujourd'hui au groupe de commercialiser sa gamme très large de médicaments adaptés aux besoins locaux.

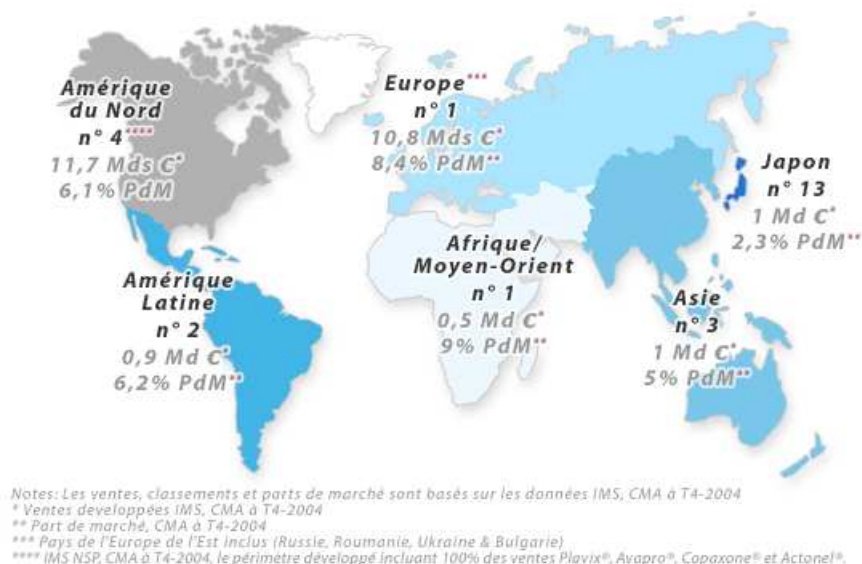


FIG. 1 – Présence internationale de Sanofi-Aventis



### 3.2 Médicaments et Produits

L'entreprise par son investissement dans la recherche et le développement élabore de nouveaux médicaments dans de nombreux axes thérapeutiques majeurs tels que :

- le cardio-vasculaire,
- la thrombose,
- l'oncologie,
- le diabète,
- le système nerveux central,
- la médecine interne et les vaccins.

Le groupe possède actuellement six blockbusters :

- Lovenox<sup>®</sup>/Clexane<sup>®</sup> (cardiovasculaire, thrombose) :

Traitement préventif des thromboses. Il est également utilisé comme traitement curatif des phlébites, et dans certains cas d'angine de poitrine et d'infarctus du myocarde.



- Plavix<sup>®</sup> (cardiovasculaire, thrombose) :

Traitement préventif des complications artérielles chez des patients ayant présenté un accident vasculaire. Il est aussi utilisé chez certains patients souffrant d'un syndrome coronaire aigu en association avec l'aspirine.



- Taxotere<sup>®</sup> (oncologie) :

Traitement du cancer du sein et du cancer du poumon.



- Allegra<sup>®</sup>/Telfast<sup>®</sup> (médecine interne) :

Traitement de la rhinite allergique saisonnière ou de l'urticaire idiopathique cutanée.



- Stilnox<sup>®</sup>/Ambien<sup>®</sup>/Myslee<sup>®</sup> (système nerveux central) :

Traitement de l'insomnie occasionnelle, transitoire ou chronique.



- Eloxatine<sup>®</sup> (oncologie) :

Anticancéreux cytotoxique utilisé dans le traitement du cancer colorectal.



Outre ces produits phares, Sanofi-Aventis possède également d'autres médicaments de référence tels que Delix<sup>®</sup>/Tritace<sup>®</sup>/Triatec<sup>®</sup> (traitement de



l'hypertension artérielle et de l'insuffisance cardiaque), Lantus<sup>®</sup> (traitement du diabète), Aprovel<sup>®</sup>/Avapro<sup>®</sup> (traitement de l'hypertension artérielle), et Copaxone<sup>®</sup> (traitement de la sclérose en plaques).

### 3.3 Recherche et Développement

Avec un budget de près de 4 milliards d'euros en 2004, Sanofi-Aventis donne les moyens correspondants à son axe stratégique prioritaire : la recherche. Aujourd'hui, la recherche du groupe possède l'un des plus riches et innovants portefeuilles de toute l'industrie pharmaceutique mondiale avec 128 molécules en développement dont 48 en phases avancées (phase II et III). Présent sur 3 continents, 11 500 chercheurs travaillent dans plus de 20 sites.

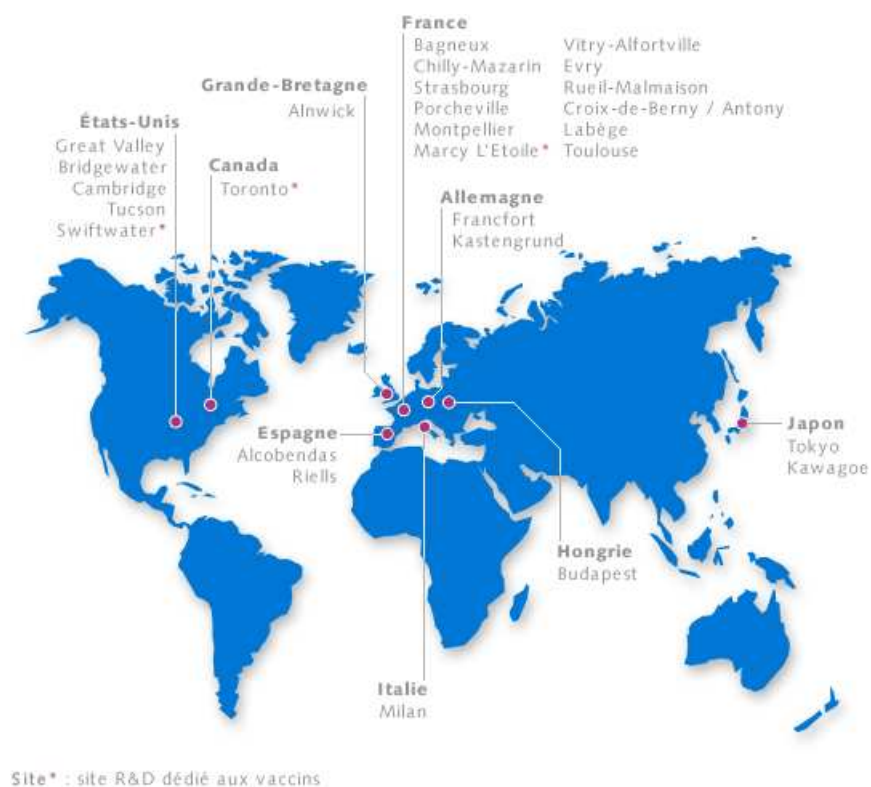


FIG. 2 – Recherche pharmaceutique et vaccins

### 3.4 Le département GMPK

Le département GMPK (*Global Drug Metabolism and Pharmacokinetics*) à Francfort sur le Main est dirigé par le Docteur Jochen Maas.



### 3.4.1 Organistaion

Ce département est subdivisé en 5 services :

- eADME/Métabolisme,
- Pharmacocinétique in vivo / NMR Technologies,
- Bioanalyses,
- Pharmacocinétique Clinique,
- Pharmacocinétique des Populations.

Mon stage de Master 1 s'est déroulé dans le service de Pharmacocinétique/Pharmacodynamie des populations.

Cette section procède au traitement des données pharmacocinétiques et pharmacodynamiques des essais cliniques. Des méthodes mathématiques et statistiques sont utilisées afin d'obtenir une idée précise de la posologie d'un médicament destiné à une population.

### 3.4.2 Environnement humain

Le service de pharmacocinétique/pharmacodynamie des populations compte 11 personnes répartis sur 5 sites : Bridgewater(Etats-Unis), Great Valley (Etats-Unis), Francfort (Allemagne), Paris (France), Montpellier(France).

Le groupe de Francfort est constitué des membres suivant :

- **Willi Weber**, médecin et chimiste, tient les fonctions de pharmacométricien et de chef d'équipe,
- **Diether Rueppel**, biophysicien, occupe le poste de pharmacométricien,
- **Heiner Speth**, informaticien.

L'équipe de Francfort travaille en collaboration avec l'Université de Berlin (Allemagne).

### 3.4.3 Environnement technique

Le service de pharmacocinétique/pharmacodynamie des populations travaille avec le système d'exploitation Linux. La distribution utilisée est la Suse 8.0. Une dizaine de machines constitue le parc informatique de ce service.

L'équipe travaille essentiellement avec le programme de régression mathématique NONMEM codé en FORTRAN. Ce dernier signifie "Non Linear Mixed Effect Modeling". Il a été développé par S.Beal et L.Sheiner, tout deux membres du NONMEN Project Group de l'Université de Californie à San Francisco.

S-plus est un langage de programmation statistique et graphique, interprété



et orienté objet. Il permet l'analyse statistique des résultats obtenus à partir de NONMEM. Il est utilisé en particulier dans le domaine de la recherche et de l'enseignement.

Afin de ressortir les différents graphiques et résultats sous forme de rapport, l'équipe de Francfort utilise le langage de mise en page professionnelle Latex. Ce rapport est un exemple de ce que peut rendre un document généré par Latex.



## 4 La pharmacocinétique/pharmacodynamie des populations

Une partie du développement d'un médicament passe par l'analyse des relations existantes entre la dose administrée et les effets thérapeutiques et/ou indésirables entraînés. Cette analyse repose d'une part sur l'étude de la pharmacocinétique (PK) et de la pharmacodynamie (PD) de la molécule et d'une autre part sur l'existence de relations entre ces deux entités (PK-PD). La mission du service dans lequel j'ai été intégré est de réaliser cette analyse afin d'en ressortir des prévisions utiles à des études ultérieures ou de détecter des éventuelles erreurs provenant de certains patients.

### 4.1 La pharmacocinétique

La pharmacocinétique correspond à l'étude du devenir du médicament dans l'organisme. Elle permet d'adapter ainsi la posologie et la voie d'administration. L'étude est fondée sur les variations de concentration plasmatique du médicament. En effet, ceci est le seul paramètre pharmacocinétique individuel directement mesurable. Ces variations traduisent les différentes étapes du médicament dans l'organisme :

- l'absorption : processus par lequel le médicament passe de son site d'administration à la circulation générale,
- la diffusion : distribution de la molécule dans les différents tissus de l'organisme dépendant des caractères physico-chimiques du médicament,
- le métabolisme : transformation par réaction enzymatique du médicament aboutissant à la formation de plusieurs composés actifs ou inactifs,
- l'élimination : excrétion du médicament hors de l'organisme.

### 4.2 La pharmacodynamie

La pharmacodynamie permet de déterminer la relation dose-efficacité du médicament. En effet, elle essaie de quantifier les effets thérapeutiques ou indésirables en fonction des doses administrées. De nombreux paramètres interviennent et rendent cette relation difficile et complexe. L'objectif est d'obtenir une posologie avec un effet thérapeutique optimum et des effets indésirables, synonyme de toxicité, minimisés.



### 4.3 Les paramètres utilisés

L'objectif principal du département PKPD est d'étudier la variabilité entre les individus d'une population et sa relation avec les résultats de pharmacocinétique et de pharmacodynamie. Dans les études de PKPD, les paramètres pharmacocinétiques utilisés tels que la clairance ou le volume de distribution sont définis à l'aide d'autres paramètres expliqués ci-dessous.

– **Paramètres à effet fixe :**

Portant le nom de  $\theta$ , ces paramètres rassemblent l'ensemble des effets fixes provoquant des variations dans une population sur un paramètre pharmacocinétique donné. Ces effets fixes peuvent avoir comme covariables : le sexe, la masse corporelle, l'âge ou encore une maladie particulière. Les  $\theta$  permettent de définir la valeur moyenne d'un paramètre de population (ex : clairance) et/ou la relation moyenne mesurable entre ce paramètre et une covariable (ex : relation entre la clairance et la masse corporelle). A partir de la définition de ces paramètres de population, les dosages peuvent être adaptés en fonction des caractéristiques des individus de la population.

– **Paramètres relatifs à la variabilité interindividuelle :**

Ces paramètres nommés  $\eta$  permettent de quantifier la variabilité existante entre les différents individus d'une population. En effet, pour un individu donné, la valeur d'un de ses variables peut différer par rapport à la valeur obtenue à partir des paramètres à effet fixe d'une population. Cette variabilité interindividuelle peut être considérée comme un paramètre à effet aléatoire ayant zéro de moyenne.

– **Paramètres relatifs à la variabilité intra-individuelle :**

Ces paramètres appelés  $\sigma$  traduisent l'existence d'une variabilité résiduelle, aléatoire et non explicable.

L'exemple ci-dessous montre l'application des deux premières catégories de paramètres.

$$Cl_i = \theta_1 \times \frac{WT}{70Kg} + \eta_i$$

L'équation exprime la clairance d'un individu ( $Cl_i$ ). Le paramètre  $\theta_1$  est un paramètre à effet fixe qui représente la valeur moyenne de la clairance dans la population étudiée. Le terme  $WT$  correspond à la covariable de la masse corporelle de l'individu. Cette valeur est divisée par la masse corporelle



d'un individu moyen. Enfin, le paramètre  $\eta_i$  représente la variabilité intra-individuelle traduisant le fait que la clairance d'un individu n'est jamais calculable grâce aux seuls paramètres fixes de population.

#### 4.4 Définition du modèle

Tout modèle en pharmacocinétique / pharmacodynamie peut être défini par les paramètres vus précédemment. Un modèle représente l'individu comme un volume pour lequel il existe une entrée et une sortie.

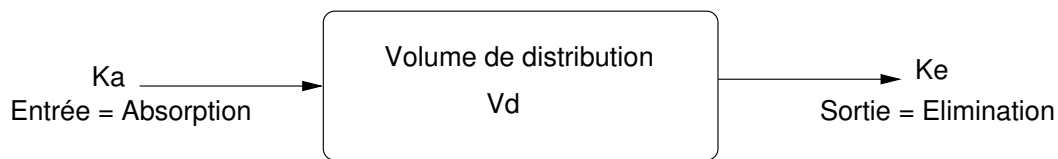


FIG. 3 – Principe du modèle à compartiment

Ce volume peut être constitué d'un ou de plusieurs compartiments. Le nombre de compartiment dépend du mode d'administration du médicament, de sa diffusion dans l'organisme et de son élimination. Il est déterminé à partir de la courbe d'évolution des concentrations plasmatiques en fonction du temps. Cette courbe de forme exponentielle est transformée de manière semi-logarithmique. Le nombre de droites obtenues détermine alors le nombre de compartiment du modèle. Le modèle à deux compartiments est le modèle hypothétique le plus fréquemment utilisé en pharmacocinétique clinique. Il peut être décrit de la manière suivante.

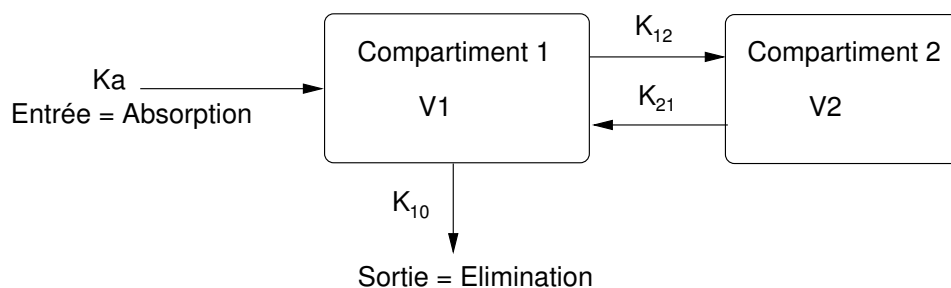


FIG. 4 – Modèle pharmacocinétique à deux compartiments

Le premier compartiment correspond au sang et aux organes bien perfusés (foie, rein, etc) alors que le second englobe les tissus moins bien irrigés.





Les échanges entre ces compartiments sont régis par les deux constantes  $k_{12}$  et  $k_{21}$ . Le nombre de compartiments reflète généralement la complexité d'un modèle. Plus le modèle possède de compartiments, plus les paramètres pharmacocinétiques à prendre en compte sont nombreux.

## 4.5 Les outils :

### 4.5.1 NONMEM

Après détermination du modèle à utiliser et de la valeur initiale des différents paramètres, le modèle est traduit en langage NONMEM et sauvegardé dans un fichier texte nommé "control file" (annexe 8.1).

D'un autre côté les données des études cliniques sont récupérées et formatées en un fichier texte utilisable pour la suite. Ce fichier contient l'ensemble des informations relatives aux patients. Il peut s'agir par exemples des doses administrées, des heures de prélèvements, des covariables (masse corporelle, âge, etc). Ce fichier texte est appelé "data file" (annexe 8.2).

Ces deux fichiers sont utilisés par NONMEM. Ce logiciel de régression mathématique est composé de trois modules :

- le programme NONMEM lui-même : programme de régression adaptable à différents types de données.
- PREDPP : puissant package de routine utilisé par NONMEM pour calculer des prévisions en terme de données pharmacocinétiques de population.
- NM-TRAN : module permettant à l'utilisateur de définir d'autres entrées nécessaires aux deux composants précédents et qui détecte un certain nombre d'erreurs que celui-ci aurait pu commettre ce faisant.

Le programme est lancé sur le modèle à partir des deux fichiers d'entrée. Celui-ci calcule les paramètres pharmacocinétiques du modèle et leur distribution statistique en boucle fermée. NONMEM commence tout d'abord par une évaluation des paramètres initiaux hypothétiques puis minimise la fonction objective. Cette dernière permet de déterminer la valeur des paramètres qui assure la plus grande probabilité d'apparition des résultats observés. Les résultats d'analyse de NONMEM sont stockés sous deux autres types de fichiers :

- un fichier report contenant les résultats d'analyses comme la valeur finale des paramètres ou encore les éventuels messages d'erreur,
- des fichiers contenant les valeurs des individus comme par exemple les variabilités interindividuelles.



Cette opération peut être répétée en faisant varier les paramètres initiaux et/ou le nombre de compartiments jusqu'à l'obtention d'un modèle jugé satisfaisant. Une fois le meilleur modèle trouvé, il peut être utilisé pour la simulation de protocoles expérimentaux et à terme, la détermination d'un nouveau protocole expérimental optimisé.

#### 4.5.2 S-PLUS

S-plus est en fait la version commercialisée du langage S (Insightful Corporation). Cette version possède une grande variété de fonctions statistiques et un module graphique important. A partir de la manipulation d'objet (matrices, vecteurs, tableaux de données), ce langage permet de traiter et de ressortir les données sous différentes formes : histogramme, diagramme, graphes, etc. De plus, S-plus est un langage mathématique permettant le calcul matriciel.

Au sein du service de pharmacocinétique/pharmacodynamie des populations, de nombreuses fonctions sont développées dans ce langage. Ces dernières permettent le lancement de méthodes d'optimisation des modèles :

- variation des valeurs initiales des thêtas,
- bootstrap,
- jackknife.

Ces trois méthodes ont été automatisées et développées par les trois stagiaires l'an dernier. Elles seront détaillées dans la suite du rapport.

A la suite de leur travail, Guillaume Gorson-Tanguy et moi-même avons participé à l'amélioration de ces fonctions ainsi qu'à l'intégration de deux nouvelles méthodes :

- variation des valeurs initiales d'un seul thêta,
- tests statistiques sur chaque individu d'un échantillon.



## 5 Stage

Ce stage s'est déroulé en deux phases principales. Tout d'abord, la première partie a consisté en la découverte du S-plus et des méthodes utilisées au sein du service. Nous avons alors pu améliorer ces méthodes suivant les besoins de l'équipe. Puis dans un second temps, nous avons participé au développement de deux nouvelles fonctions.

### 5.1 Présentation des trois méthodes existantes :

#### 5.1.1 Méthode Initial Thetas

Le but ici est de déterminer le meilleur modèle pharmacocinétique à effets mixtes pour une étude donnée. Ce dernier est caractérisé par un nombre de compartiments et par la valeur des paramètres pharmacocinétiques. Comme nous l'avons vu précédemment les paramètres sont exprimés par l'intermédiaire d'un effet fixe constant dans la population et par un effet aléatoire (variabilité inter et intra-individuelle). Le  $\theta$  correspond à l'effet fixe. Une initialisation de chacun des  $\theta$  dans le fichier de contrôle est nécessaire avant de lancer le programme NONMEM. La variation des  $\theta$  peut entraîner alors une variation de la fonction objective. On peut ainsi obtenir une fonction objective plus faible traduisant un meilleur modèle.

Cette variation automatique des  $\theta$  a été réalisée par Justine Lahaye l'an passé. Grâce à la fonction `initial.thetas`, la valeur des différents lots de  $\theta$  est modifiée. Pour chaque lot de valeur, NONMEM est lancé. On obtient ainsi au final un tableau avec pour chaque valeur de  $\theta$ , la fonction objective retournée.

#### 5.1.2 Méthode Bootstrap

Cette méthode permet l'étude de la représentativité de l'échantillon. Inventée par Bradley Efron, elle repose sur la simulation de données à partir d'un nombre limité d'observations. Il s'agit d'une méthode de rééchantillonnage classique qui permet de tester le modèle avec de nouveaux sets de données et d'évaluer la confiance que l'on peut avoir en l'échantillon original, à savoir dans notre cas si celui-ci est représentatif de la population à laquelle le médicament s'adresse.

Un échantillon bootstrap doit contenir autant de valeurs que l'échantillon original. Pour le créer, il suffit de faire un tirage aléatoire avec remise parmi l'échantillon original. De ce fait si l'échantillon original contient  $N$  individus,



le tirage aléatoire avec remise donnera au maximum la création de  $\frac{(2N-1)!}{[N! \times (N-1)!]}$  échantillons de composition différente. Prenons l'exemple d'un échantillon composé de trois personnages tux (très connus sous Linux). Voici ce que donne une génération de nouveaux échantillons avec la méthode Bootstrap.

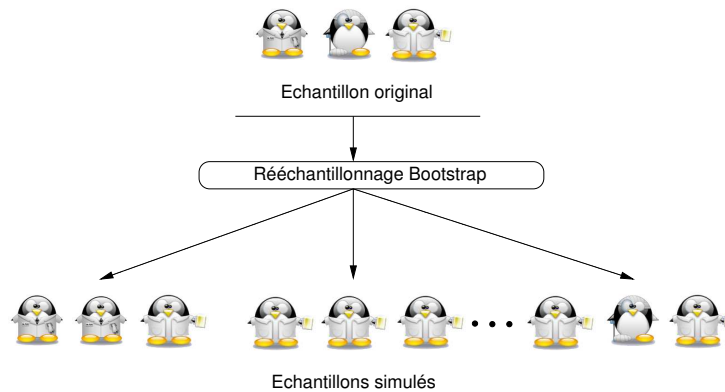


FIG. 5 – Rééchantillonnage par la méthode Bootstrap

Dans notre cas, l'échantillon joue le rôle de population et l'ensemble des échantillons créés constitue une population à partir de laquelle on réalise des statistiques (médian, écart-type,...). L'échantillon d'origine peut donc être comparé à la population théorique générée par le bootstrap. Si des différences significatives sont détectées, on peut conclure que l'échantillon d'origine est biaisé et que les résultats obtenus ne concordent pas avec ceux qui auraient été obtenus à partir d'un échantillon représentatif de la population.

La génération aléatoire des nouveaux échantillons a été automatisée l'an dernier par Baptiste Moulinier. La fonction `pkpd.bootstrap` permet ainsi la création de nouveaux échantillons sur lesquels sont lancés le programme NON-MEM.

### 5.1.3 Méthode Jackknife

Cette méthode de rééchantillonnage inventée par Maurice Quenouille en 1949 permet de générer des données artificielles à partir de données d'origine sans faire intervenir la notion de hasard.

Pour générer un échantillon jackknife, il suffit de prendre l'échantillon original et d'ôter un individu.



Reprenons l'exemple avec notre échantillon de tux.

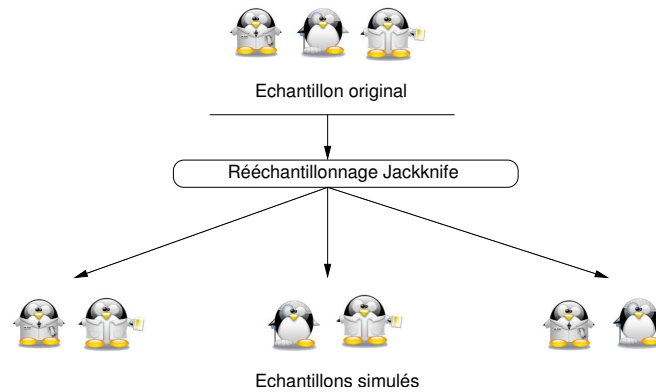


FIG. 6 – Rééchantillonnage par la méthode Jackknife

Pour un échantillon de  $N$  individus, on obtiendra donc  $N$  échantillons jackknife de taille  $N-1$  individus. Le but ici est de déterminer si un individu particulier a une influence significative sur les résultats de régression de NONMEM. En effet si un échantillon ne comprenant pas l'individu  $X$  donne des résultats de fonction objective significativement différents de ceux obtenus à partir des données originales, c'est que cet individu biaise les résultats de l'échantillon original. Celui-ci est donc isolé et des analyses médicales pourront par la suite expliquer les raisons de ce biais. Le dosage du médicament pourra alors être adapté de manière spécifique aux individus présentant les mêmes caractéristiques.

La fonction `pkpd.jackknife` créée par Matthieu Chosseler permet la génération de ces nouveaux échantillons et leur traitement par NONMEM.

## 5.2 Améliorations apportées :

Durant la première partie de notre stage au sein du département GMPK, l'équipe a souhaité réaliser des améliorations sur les méthodes existantes.

### 5.2.1 Les besoins :

NONMEM est un programme gourmand en ressource. Lorsque celui-ci est lancé, il arrive assez régulièrement que le processus dépasse l'heure de traitement et occupe la majeure partie des ressources du système. De plus il se peut que le programme "crash" et ne renvoie pas de valeur pour certains échantillons. A la suite de ceci, un rapport est édité alors que les valeurs obtenues sont nulles.



Pour l'équipe, ceci constitue un inconvénient qui est valable pour les trois fonctions développées. L'idée était donc de bloquer la génération de rapport dans le cas où un ou plusieurs des différents NONMEM lancés par une fonction n'avaient pas abouti.

### 5.2.2 Modifications des applications :

#### Analyse des flowcharts :

Le fonctionnement des différentes méthodes est similaire en plusieurs points. Nous allons donc réaliser une analyse générale de ces fonctions par l'intermédiaire d'aperçu écran.

Nous lançons pour notre exemple la méthode `initial.thetas`. Ceci peut être fait à partir du fichier control file lui même ou à partir de l'archive contenant l'ensemble des données de l'étude et le control file.

```
> initial.thetas(src=3541320, path=".", dest=".")
[1] "Initial Thetas has been started"
[1] "Main directory /work/trial/poitiers/matthieu2/input/./initval.3541320.4 created"
[1] "Archive 3541320 untared"
```

Ici, nous rentrons un numéro d'archive en paramètre (`src`) ainsi que le dossier où est présente cette archive (`path`) et le dossier de destination des résultats (`dest`). Des messages précisent que la méthode a bien démarré, que le dossier de travail a bien été créé et que l'archive est décompactée afin de récupérer les données utiles pour la suite. Dans le cas de l'utilisation d'un fichier de contrôle comme paramètre "`src`", le programme lance tout d'abord une première fois le programme NONMEM qui va alors récupérer les fichiers et en faire une archive utilisable.

```
CONTROL ARCHIVE REPORT DATA TABLE
"nm.final4.con" "3541320" "nm.final4.rep" "nmTotal.dat" "final4.TAB"
PATH
"/work/trial/poitiers/matthieu2/input/./initval.3541320.4"
[1] "Got usefull data from original report and control files"
```

Les différents fichiers décompactés apparaissent ainsi que le nouveau "`path`". Le programme récupère les données des fichiers dans des variables pouvant être ensuite traitées.

```
[1] "Calculating the new thetas"
[1] "New thetas calculated"
```

Dans l'exemple de la fonction `initial.thetas`, le programme lance le calcul de nouveaux thêtas et crée alors un nouveau fichier de contrôle pour chaque



nouveau lot de valeurs. Dans le cas des fonctions jackknife et bootstrap, le traitement se fait sur le fichier data. Jackknife recrée des nouveaux fichiers data en enlevant un individu. Bootstrap lui génère des fichiers data avec des individus choisis de manière aléatoire. Pour chaque nouveau fichier, un sous répertoire est créé.

```
[1] "Starting the 25 nmr runs"
[1] "sleep 1 Sec. Nmr run number 1 started"
[1] "sleep 1 Sec. Nmr run number 2 started"
[1] "sleep 1 Sec. Nmr run number 3 started"
[1] "All the nmr runs have been started"
```

Le nombre de nouveaux sous-répertoires est par défaut de 25 mais il peut être modifié lors du lancement de la fonction par le paramètre N. Ici, NONMEM est lancé dans chaque sous-répertoire contenant les nouveaux fichiers control ou data.

```
[1] "Nmr run number 1 finished ( 24 left )"
[1] "Nmr run number 2 finished ( 23 left )"
[1] "Nmr run number 3 finished ( 22 left )"
[1] "Nmr run number 4 finished ( 21 left )"
[1] "Nmr run number 5 finished ( 20 left )"
```

A la suite, une boucle d'attente permet de suivre l'avancée des différents programmes NONMEM lancés et d'éviter la création du rapport tant que tous les résultats ne sont pas obtenus.

```
[1] "Excel report created"
[1] "Latex report created"
[1] "Information about the different reports : "
[1] "Type soffice /work/trial/poitiers/matthieu2/input/initval.3541320.6/doc/excel/initval.3541320.6.xls in a shell to read the excel report 'initval.3541320.6.xls'"
[1] "Type xdvi /work/trial/poitiers/matthieu2/input/initval.3541320.6/doc/latex/initval.3541320.6.dvi in a shell to read the latex report 'initval.3541320.6.dvi'"
[1] "You can also draw plots with the new pharmacokinetic models by using the function create.plots."
[1] "To get the plots for all of the models, use the following command : create.plots(src=/work/trial/poitiers/matthieu2/input/./initval.3541320.6) where src is a CHARACTER STRING (write with quotation marks)."
[1] "Otherwise use the parameter 'arch' as a vector of the numbers of the archives"
[1] "Initial Thetas program finished"
```

Une fois les différents programmes NONMEM terminés, un rapport sous deux formats (.xls et .dvi) est généré à partir des différents résultats récupérés dans chaque sous-répertoire.

Ce flowchart est résumé de manière graphique dans la figure suivante.



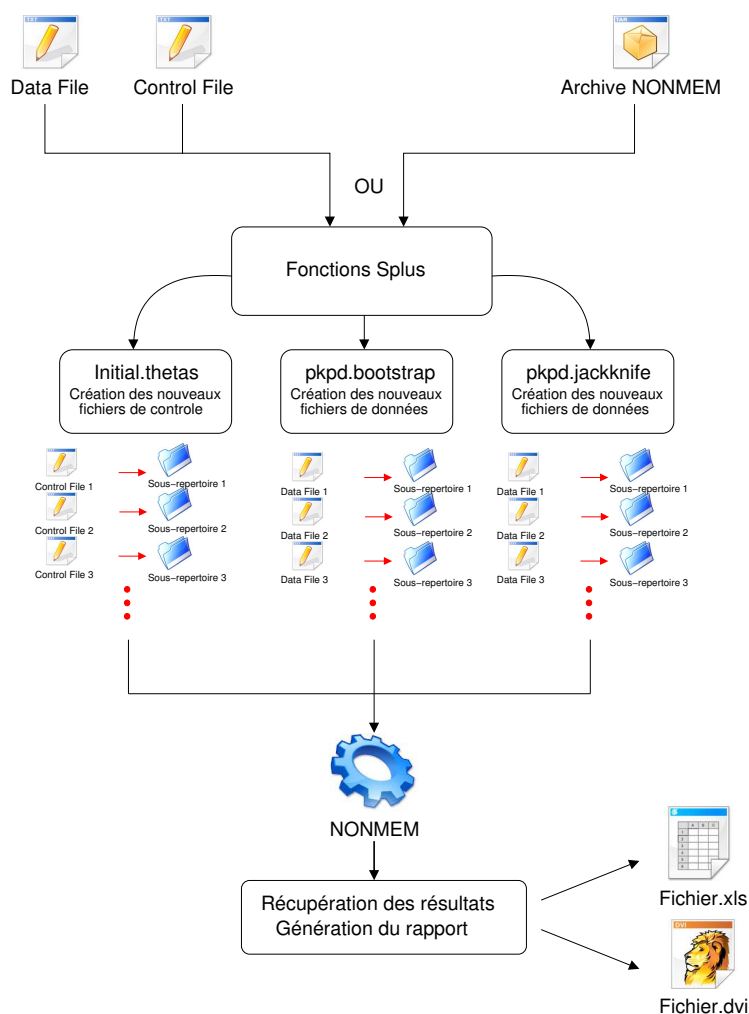


FIG. 7 – Flowchart général

### Optimisation des fonctions :

Afin d'éviter la création de rapport dans le cas d'un plantage d'un des NONMEM lancés, nous avons tout d'abord procédé au découpage des fonctions. Pour cela, un nouveau paramètre a été créé nommé "runlevel". Celui-ci peut posséder trois valeurs :

- 0 : c'est le mode par défaut, la fonction se déroule comme nous l'avons vu précédemment avec une création de rapport à la suite des programmes NONMEM,
- 1 : les programmes NONMEM sont lancés et la fonction s'arrête,
- 2 : ce mode lance exclusivement la création du rapport.





Afin de permettre la création du rapport en runlevel 2, la fonction a besoin de toutes les données relatives au runlevel 1 (le répertoire source path, le répertoire de destination dest, le nombre de NONMEM lancés, le numéro d'archive arch.num, etc). Un objet de type "list" a donc été déclaré afin de stocker ces différentes informations et d'assurer la liaison entre les deux modes. Avant de lancer la création de rapport, le programme doit s'assurer que tous les programmes NONMEM lancés ont bien aboutis sans problème particulier. En effet l'ancienne version de la fonction ne prenait pas en compte les plantages possibles du logiciel et générait alors un rapport avec des valeurs nulles. Pour résoudre ce problème, une nouvelle sous-fonction nommée "testing.job" a été développée. L'absence d'un certain fichier dans le sous répertoire, indique que NONMEM n'a pas fonctionné. La sous-fonction vérifie donc dans chaque sous-répertoire la présence ou non de ce fichier et renvoie alors une liste de boolean. Dans la fonction principale, cette liste est analysée et dans le cas où un zéro est présent (synonyme de l'absence d'un fichier dans un sous-répertoire et donc du plantage de NONMEM), la fonction s'arrête en renvoyant le message suivant :

```
Problem in initial.thetas(list.param = list4, runlevel = 2): Error: NMR Job Failed  
Use traceback() to see the call stack
```

Cette sous-fonction est exécutée dans le runlevel 2 de la fonction mère. Dans le cas d'un plantage de NONMEM, la suite du programme est interrompue ce qui évite la création d'un rapport contenant des valeurs nulles.

Ce premier travail nous a permis de comprendre les différentes méthodes statistiques utilisées ainsi que d'apprendre le langage Splus, jusqu'alors inconnu pour nous.

### 5.3 Nouvelles fonctions développées :

Dans un deuxième temps, l'équipe nous a soumis l'idée de deux nouvelles fonctions. La première aurait pour but d'observer les différences de résultats avec la variation d'un seul des thêtas dans le control file. Concernant la seconde Simplefit, cette fonction permettrait l'analyse sur chaque individu d'un échantillon et non plus sur l'échantillon en entier.



### 5.3.1 Fonction Variation Onetheta

#### Analyse de l'existant :

Les effets fixes des paramètres pharmacocinétiques sont traduits comme nous l'avons vu précédemment par les paramètres thêtas. Ceux ci sont définis dans le fichier control file et plus précisément dans le bloc \$THETA.

```
$THETA (0.0001,0.664,130)|  
$THETA 0.09016 FIX  
$THETA (0,5.87,100)  
$THETA (0.0001,4.63,120)  
$THETA (0.0001,86.6,1000)  
$THETA (0.0001,9.26,1000)  
$THETA (1,13.4,1000)  
$THETA (0.0001,0.959,1000)  
$THETA (0.0001,2.8,1000)  
$THETA 0 FIX  
$THETA (0.0001,0.793,1)
```

FIG. 8 – Bloc \$THETA extrait du control file.

La valeur centrale correspond à la valeur initiale du thêta. Celle-ci est encadrée par deux autres valeurs correspondant aux bornes inférieures et supérieures de variation réalisée par NONMEM. Un thêta peut aussi avoir une valeur fixée qui ne varie pas lors du calcul des nouveaux thêtas. Ceci est traduit dans le fichier de contrôle par une unique valeur suivie du paramètre "FIX".

L'objectif pour cette nouvelle fonction est de permettre une variation contrôlée d'un seul thêta dans la liste. Nous avons donc proposer de construire une sous-fonction permettant de créer de nouveaux fichiers de contrôle présentant une valeur de thêta fixée différente pour le thêta étudié.

#### Réalisation :

Cette méthode correspond à une nouvelle méthode d'analyse nommée variation.onetheta. Cette fonction présente la même structure que initial.thetas. La différence vient dans l'appel de nouvelles sous-fonctions.

- Calcul des nouvelles valeurs de thêta :

L'équipe souhaite pouvoir contrôler la variation du thêta en donnant comme argument le pourcentage de variation ainsi que le nombre de nouvelles valeurs à calculer. Ces deux nouveaux paramètres ont donc été



introduits respectivement sous le nom de “variation” et “N.onethetas”. Le  $\theta$  à faire varier est indiqué par sa position dans la liste du fichier de contrôle. A partir de la valeur initiale donnée dans ce fichier, la fonction “onetheta.sampling” calcule la liste des nouvelles valeurs selon les deux arguments vus précédemment.

- Création des nouveaux fichiers de contrôle :  
Chaque nouvelle valeur obtenue est ensuite insérée dans un nouveau fichier de contrôle pour le  $\theta$  correspondant. Cette valeur est fixée afin de ne pas être modifiée par NONMEM. L’opération est réalisée par la fonction “onetheta.drawing”. Pour chaque nouvelle valeur de  $\theta$ , un sous-répertoire est créé dans lequel est placé le nouveau fichier de contrôle ainsi qu’une copie du fichier de données original.
- Lancement des programmes NONMEM :  
De la même manière que pour les méthodes pré-existantes, NONMEM est lancé dans chaque sous-répertoire.



Ces trois premières étapes sont résumées dans le figure suivante :

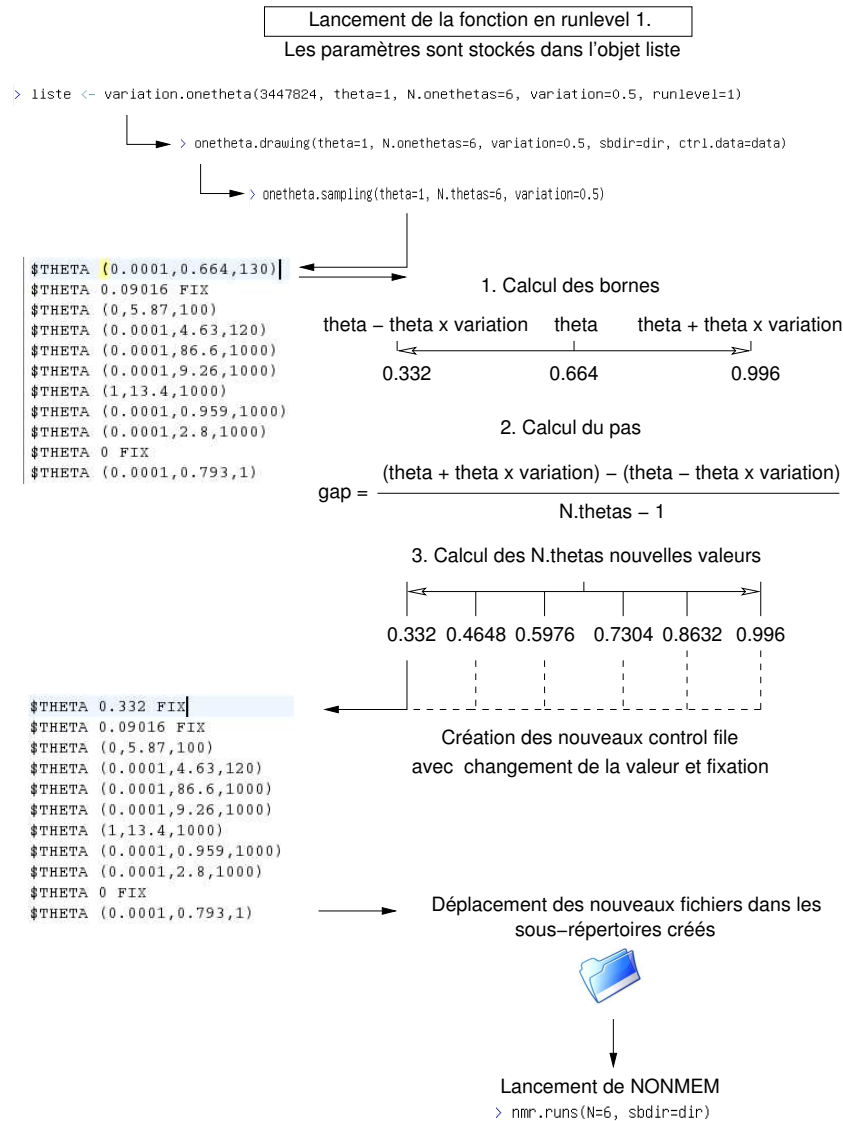


FIG. 9 – Flowchart Variation Onetheta (mode 1)

– Récupération des données et analyse :

La fonction en runlevel 2 permet ensuite le rassemblement des résultats. Le statut de chaque NONMEM lancé est récupéré avec la fonction “get.of.onetheta”. NONMEM procède pour chaque lot de données à une minimisation de la fonction objective. Si celle-ci n’aboutit pas, le programme renvoie le statut “MINIMIZATION TERMINATED”. Selon les



besoins de l'équipe, une sous-fonction "test.terminated.onetheta" a été développée afin de lancer automatiquement dans ce cas la méthode initial.thetas. Cette sous-fonction prend en paramètre le résultat de "get.of.onetheta". Dans le cas d'une minimisation terminée, le recalcul de la fonction objective avec 5 nouveaux lots de thêtas est lancé. "test.terminated.onetheta" récupère la valeur de fonction objective pour chaque lot et isole la plus faible. Celle-ci vient ensuite remplacée celle obtenue avec le statut "MINIMIZATION TERMINATED".

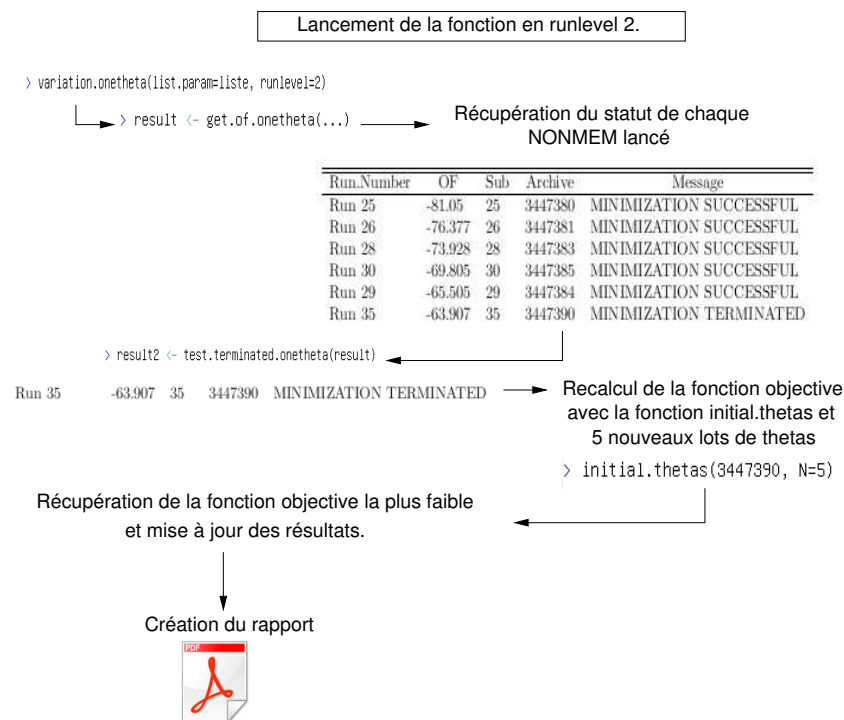


FIG. 10 – Flowchart Variation Onetheta (mode 2)

– Création du rapport :

Une fois les données vérifiées et mises à jour, le rapport est généré par une nouvelle fonction "create.reports.onetheta". Celle-ci crée un rapport spécifique à la méthode variation.onetheta. Elle permet par l'intermédiaire du code Latex de ressortir le rapport sous format PDF. Ce dernier contient en plus du tableau de résultats, un graphique représentant la variation de fonction objective en fonction des différentes valeurs de thêtas calculées. L'équipe a demandé aussi l'apparition sur ce graphique de différentes valeurs de  $\chi^2$  pour un risque de première espèce de 0,05



, 0,01 et 0,001. Les lignes représentant ces valeurs sont affichées seulement dans le cas où la variation de la fonction objective est inférieure. Un exemple de graphique vous est présenté à la suite.

Figure 2: Variation of Theta CL Theta (OF-initialOF)

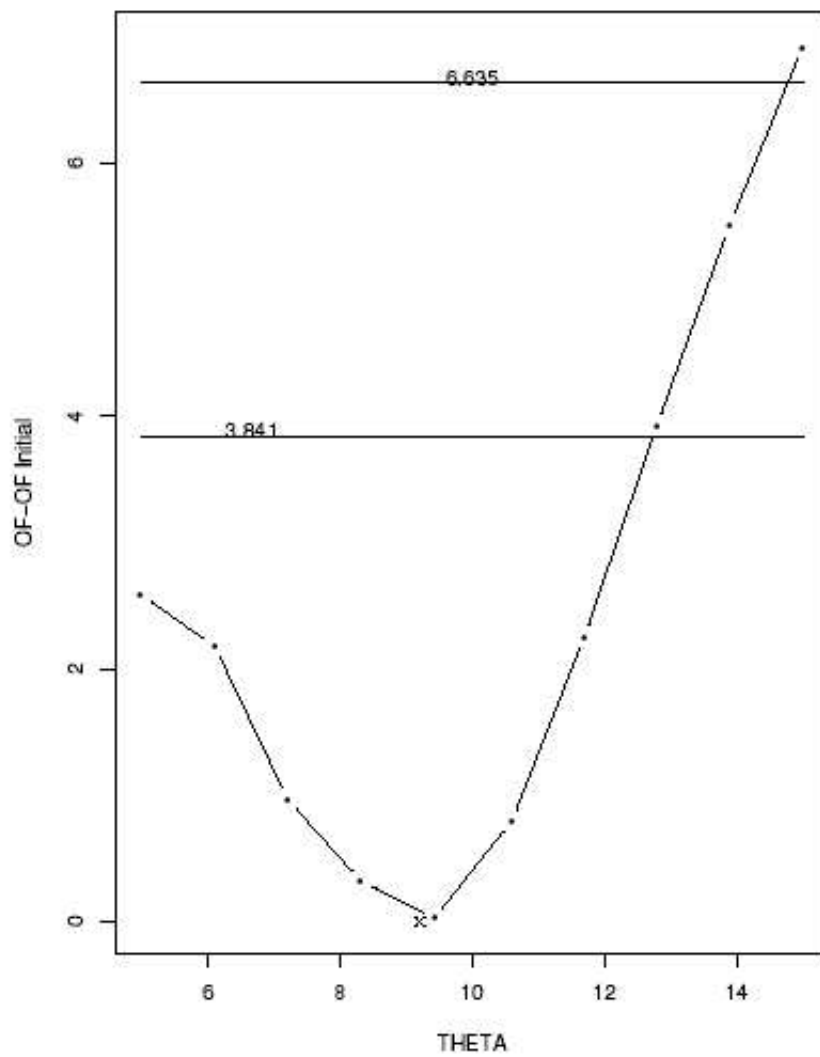


FIG. 11 – Exemple de graphique généré



### 5.3.2 Fonction Simplefit

L'objectif de cette fonction est de permettre une étude statistique sur chaque individu d'une population et non plus sur la population entière. Ainsi l'individu peut être comparé à l'ensemble de la population et à chaque individu de cette population.

#### **Analyse de l'existant :**

Les données concernant les patients sont stockées comme nous avons pu le voir précédemment dans le fichier data. Ces individus sont identifiés par un numéro ID. Le but de la nouvelle fonction va donc être de créer de nouveaux fichiers contenant les données propres à chaque individu de l'échantillon.

#### **Réalisation :**

Cette nouvelle fonction "simplefit" est basée sur la fonction "pkpd.jackknife". La différence vient dans la création des nouveaux fichiers de données par la sous-fonction "sf.sampling".

- Création des nouveaux fichiers de données :

La sous-fonction "sf.sampling" analyse le fichier de données original afin de dénombrer le nombre d'individu et de stocker leur identifiant. A partir de ces deux paramètres, de nouveaux fichiers de données sont créés. Ils contiennent chacun les données relatives à un individu. Ces derniers sont placés dans les sous-répertoires correspondant. Pour lancer une étude individuelle avec "simplefit", le fichier de contrôle doit être modifié. Ces modifications se font de façon manuelle. Le fichier modifié doit être donné comme argument en complément de l'ancien fichier de contrôle ou de l'archive. Il peut aussi être modifié directement dans l'archive. Dans ce cas la fonction ne prendra comme source que l'archive modifiée. Le flowchart de la fonction en mode 1 vous est présentée.



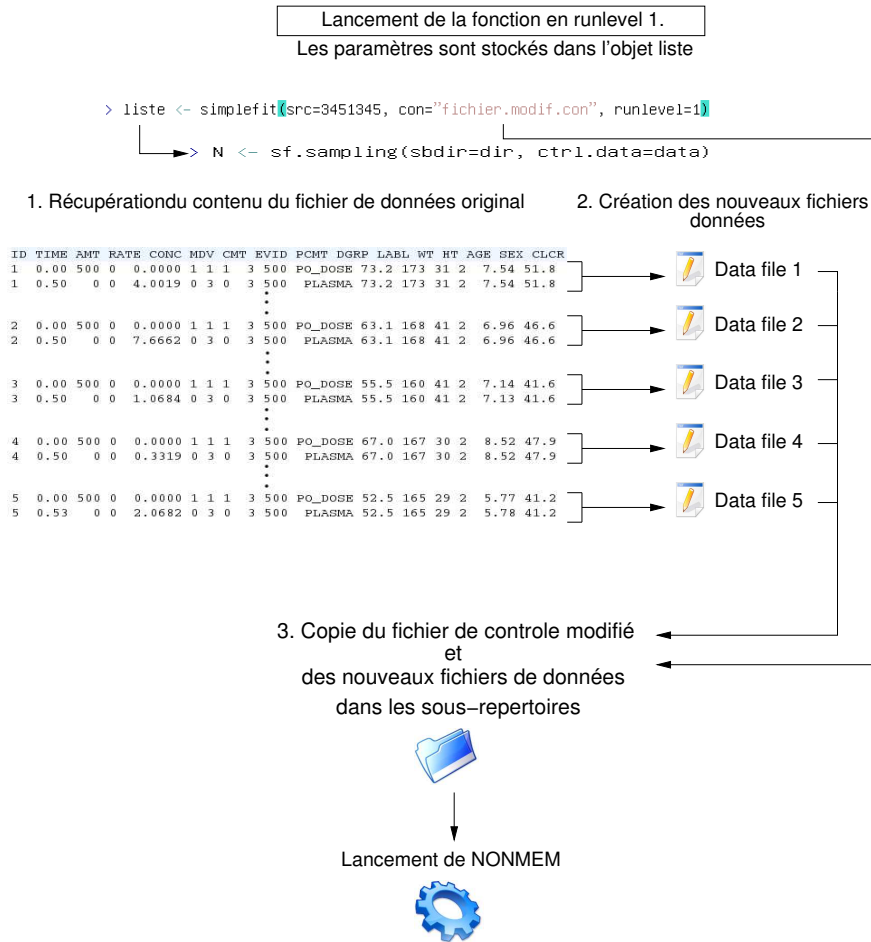


FIG. 12 – Flowchart Simplefit en mode 1

- Lancement de NONMEM et création des rapports :  
La fonction “sf.sampling” renvoie comme résultat à la fonction “simplefit” le nombre d’individu dans la population. Cet argument est utilisé par la suite afin de lancer les programmes NONMEM dans les différents sous-répertoires. Les résultats sont ensuite recueillis et analysés comme pour la fonction variation.onetheta. Dans le cas d’une minimisation terminée, la fonction objective est recalculée et mise à jour. Les données sont ensuite utilisées afin de produire le rapport de résultats. Pour cette méthode, le dernier module de création de rapport a été modifié afin de permettre l’inclusion d’une nouvelle colonne dans le tableau de résultats final. Cette colonne correspond au numéro d’identifiant des différents sujets de l’étude. Voici un exemple de tableau généré avec la





colonne “Subject” correspondant au numéro de patient.

Table 26: New Result

Run.Number	Subject	OF	Sub	Archive	Message
Run 9	48	-152.519	36	3451354	MINIMIZATION TERMINATED
Run 5	42	-147.827	32	3451350	MINIMIZATION SUCCESSFUL
Run 1	1	-146.657	1	3451346	MINIMIZATION SUCCESSFUL
Run 6	43	-144.9	33	3451351	MINIMIZATION SUCCESSFUL
Run 4	41	-143.171	31	3451349	MINIMIZATION SUCCESSFUL
Run 2	16	-141.167	12	3451347	MINIMIZATION SUCCESSFUL
Run 3	31	-140.698	23	3451348	MINIMIZATION SUCCESSFUL
Run 8	47	-132.465	35	3451353	MINIMIZATION SUCCESSFUL
Run 10	3	-127.773	2	3451355	MINIMIZATION SUCCESSFUL
Run 23	22	-124.335	16	3451368	MINIMIZATION SUCCESSFUL
Run 11	4	-122.046	3	3451356	MINIMIZATION SUCCESSFUL
Run 7	45	-119.941	34	3451352	MINIMIZATION SUCCESSFUL
Run 12	5	-110.067	4	3451357	MINIMIZATION SUCCESSFUL
Run 14	8	-108.471	6	3451359	MINIMIZATION SUCCESSFUL
Run 16	10	-106.156	8	3451361	MINIMIZATION SUCCESSFUL
Run 18	12	-105.378	10	3451363	MINIMIZATION SUCCESSFUL
Run 13	6	-103.146	5	3451358	MINIMIZATION SUCCESSFUL
Run 15	9	-101.405	7	3451360	MINIMIZATION SUCCESSFUL
Run 17	11	-97.63	9	3451362	MINIMIZATION SUCCESSFUL
Run 21	20	-93.595	14	3451366	MINIMIZATION SUCCESSFUL
Run 24	23	-93.514	17	3451369	MINIMIZATION SUCCESSFUL
Run 22	21	-93.361	15	3451367	MINIMIZATION SUCCESSFUL
Run 19	15	-90.577	11	3451364	MINIMIZATION SUCCESSFUL
Run 20	17	-85.439	13	3451365	MINIMIZATION SUCCESSFUL
Run 27	27	-85.178	20	3451372	MINIMIZATION SUCCESSFUL

FIG. 13 – Nouveau tableau généré

### 5.3.3 Génération des fichiers d’aide

Cette dernière partie est primordiale afin de faciliter l’utilisation des différentes fonctions créées. Durant notre stage, nous avons pris soin tout d’abord de documenter au maximum notre code. Ceci avait été entrepris par Justine, Baptiste, Matthieu et nous a permis de comprendre beaucoup plus rapidement les fonctions et le code utilisé. En complément, S-plus permet la génération de fichiers d’aide formatés. En effet par la simple entrée de deux commandes, un fichier .sgml est créé avec les champs à compléter.



```

<!doctype s-function-doc system "s-function-doc.dtd" [
<!entity % S-OLD "INCLUDE">
]
>
<s-function-doc>
<s-topics>
<s-topic> simplefit </s-topic>
</s-topics>
<s-title>
Simplefit Program
</s-title>
<s-description>
Generation of new samples from initial sample for individual studies.
and creates reports to analyze the results.
</s-description>
<s-usage>
<s-old-style-usage>
simplefit(src, con=NULL, path=".", dest=".", exp.bool=F, runlevel=0,
list.param=list(), nmrState=c())
</s-old-style-usage>
</s-usage>
<s-args-required>
</s-args-required>
<s-args-optional>
<s-arg name=" src ">
The source giving the original data to use in order to run
the program. <br>
This source can be :<br> - an archive (the src parameter is a numeric
containing the number of the archive)<br>
- a control file (the src parameter is a string containing the name
of the control file)<br>
</s-arg>
<s-arg name=" con ">
The con file is a control file modified for the individual study.
</s-arg>
<s-arg name=" path ">
The path to the src. It is the current directory by default..
If the src is an archive, it is the path to the "nmqs" directory
containing the archive, else it is the path to the control file.
</s-arg>
<s-arg name=" dest ">
The dest is the directory of destination which the nmr runs are started.
</s-arg>
<s-arg name=" exp.bool ">
A boolean saying if the user would like to have reports with logarithmic thetas.
Data converted into real parameters data (TRUE) or not (FALSE)
</s-arg>

```

FIG. 14 – Exemple de fichier .sgml.

Une fois complété, ce fichier est utilisé afin de générer 3 nouveaux types de fichiers : html, xml et sgm. Le dernier permet d'avoir l'aide directement dans la console. Par l'intermédiaire de ces nouveaux fichiers, l'utilisateur peut alors consulter très facilement l'aide de chaque fonction et sous-fonction.



## 6 Conclusion

Au cours de ce stage réalisé au sein du service Pharmacocinétique / Pharmacodynamie des populations dans le département Drug Metabolism and Pharmacokinetics de Francfort sur le Main, j'ai pu mettre en pratique ma double compétence biologie-informatique.

Avec le concours d'un autre stagiaire de l'I.U.P. "Génie Physiologique et Informatique" il nous a été demandé d'optimiser des fonctions existantes permettant le lancement de méthodes statistiques. De plus nous avons développé deux nouvelles fonctions permettant d'étudier la variation d'un paramètre initial particulier ainsi que l'analyse de chaque individu d'un échantillon. Ces méthodes sont aujourd'hui implémentées et documentées afin de permettre une utilisation optimale.

Ce stage m'a permis d'apprendre de nouveaux langages de programmation tels que le Latex ou encore le S-plus. Ce dernier langage est basé sur le langage open-source "R". Par ce stage, j'ai surtout pris conscience de l'intérêt que je portais au domaine de l'open-source et au fait que Linux pouvait correspondre aux besoins d'une entreprise de cette taille. Stabilité, Sécurité et Coût sont les critères que prône Sanofi-Aventis ici en Allemagne et auxquels Linux répond tout à fait. D'un point de vue biologique, j'ai aussi découvert l'importance de l'analyse des essais cliniques pour la suite du développement du médicament. Enfin, au delà du côté technique, ce stage m'aura permis de travailler au sein d'un grand groupe mondial et de découvrir la culture allemande.



## 7 Bibliographie

**Mathematical modeling of Pharmacokinetic Data** , David W.A Bourne  
Ph.D, College of Pharmacy Sciences Center Oklahoma University, 1993.

**An introduction to mixed effect modeling** , Lewis B. Sheiner, Thad-  
deus H. Grasela, Journal of pharmacokinetics and biopharmaceutics,  
Vol 19 No. 3 (June supplement 1991).

**S-PLUS Essentials II The Command Line** , Insightful Corp, August 2002.

**Einführung in S-Plus** , Martin Theus, Universität Augsburg LS für Rech-  
nerorientierte Statistik und Datenanalyse.

**A guide for unwilling S User** , Patrick Burns, February 2003.



## 8 Annexes

### 8.1 Control file :

```

$PROBLEM variable input delay: bateman + 2cmt disposition
$DATA nm.test.dat IGNORE=@
$INPUT ID TIME AMT RATE CONC=DV MDV CMT EVID PCMT DGRP LABL=DROP
WT HT AGE SEX CLCR LBM
$SUBROUTINES ADVAN6 TRANS1 TOL=3
$MODEL
COMP=(DEPOT,DEFDOSE)
COMP=(TRANSIT)
COMP=(CENTRAL,DEFOBS)
COMP=(PERIPHER)
$PK
  CALLFL   = -1
  EPSIL    = 1.D-4
KIDN=CLCR/6
SIZE=LBM/50
EST=MIXEST
IF (MIXNUM.EQ.2) THEN
  FMAT = THETA(8) ; ABSORPTION RATE
  FMTT=THETA(9) ; MEAN TRANSIT to ABSORPTION
ELSE
  FMAT=1
  FMTT=1
ENDIF.....
  MAT = FMAT*THETA(1)*EXP(ETA(1)) ; ABSORPTION RATE
CV=THETA(10)
  MTT= FMTT*THETA(2)*EXP(CV*ETA(1)) ; MEAN TRANSIT to ABSORPTION
CLN=THETA(3)
CLR=THETA(4)
  CL = (CLN*SIZE+CLR *KIDN) *EXP(ETA(2)) ;LOSS RATE DRUG
V3 = THETA(5)*EXP(ETA(3)) *SIZE ;CENTRAL VOL
V4 = THETA(6)*EXP(ETA(4)) *SIZE ;PERIPHER VOL
DVZ = THETA(7)*EXP(ETA(5)) *SIZE ;VZ-VSS VOL
VSS=V3+V4
VZ= VSS+DVZ
L2=CL/VZ
NN=1
KTR=(NN+1)/MTT
K23=1/MAT
K30=CL/V3
K43=L2*(VZ-V3)/(VZ-VSS)
K34=K43*V4/V3
S3 = V3
$DES
; AMT dosing
DADT(1) = -KTR*A(1)

```



```

DADT(2) = KTR*A(1) - K23*A(2)
DADT(3) = K23*A(2) - K30*A(3) - K34*A(3) + K43*A(4)
DADT(4) = K34*A(3) - K43*A(4)
$ERROR
QQ = 1
IF (F .LT. EPSIL) THEN
  QQ = 0
ENDIF.
PF = QQ * F + (1-QQ) * EPSIL
IPRED = PF
W=IPRED
IRES = DV-IPRED
IWRES = IRES/W
Y = IPRED + W*ERR(1)
$MIX
      NSPOP=2
      P(1)=THETA(11)
      P(2)=1.-THETA(11)
$THETA (0.0001,0.664,130)
$THETA 0.09016 FIX
$THETA (0,5.87,100)
$THETA (0.0001,4.63,120)
$THETA (0.0001,86.6,1000)
$THETA (0.0001,9.26,1000)
$THETA (1,13.4,1000)
$THETA (0.0001,0.959,1000)
$THETA (0.0001,2.8,1000)
$THETA 0 FIX
$THETA (0.0001,0.793,1)
$OMEGA BLOCK(1) 0.5.
$OMEGA BLOCK(1) 0.5.
$OMEGA BLOCK(1) 0.5.
$OMEGA BLOCK(1) 0 FIX
$OMEGA BLOCK(1) 0 FIX
$SIGMA 0.04
;$SIMULATION (160196) ONLYSIMULATION SUBPROB=1
;$MSFI nm.bateman2cmt.msf
$ESTIMATION METH=1 INTER PRINT=10 NOABO MAX=999 SIG=3 POSTHOC
MSF=nm.bateman2cmt.msf
;$COV.
$TABLE CL V3 V4 DVZ K23 K34 K43 MTT MAT EST
      ETA1 ETA2 ETA3 ETA4 ETA5 ;ETA6
      IPRED IRES IWRES
NOPRI ONEHEADER FILE=nm.bateman2cmt.tab

```



## 8.2 Data file :

ID	TIME	AMT	RATE	CONC	MDV	CMT	EVID	PCMT	DGRP	LABL	WT	HT	AGE	SEX	CLCR	
1	0.00	500	0	0.0000	1	1	1	3	500	PO_DOSE	73.2	173	31	2	7.54	51.8
1	0.50	0	0	4.0019	0	3	0	3	500	PLASMA	73.2	173	31	2	7.54	51.8
1	1.00	0	0	5.5022	0	3	0	3	500	PLASMA	73.2	173	31	2	7.53	51.8
1	2.00	0	0	4.8827	0	3	0	3	500	PLASMA	73.2	173	31	2	7.52	51.8
1	4.00	0	0	3.8117	0	3	0	3	500	PLASMA	73.2	173	31	2	7.49	51.8
1	6.00	0	0	2.9267	0	3	0	3	500	PLASMA	73.2	173	31	2	7.46	51.8
1	8.10	0	0	2.2403	0	3	0	3	500	PLASMA	73.2	173	31	2	7.43	51.8
1	12.00	0	0	1.2171	0	3	0	3	500	PLASMA	73.2	173	31	2	7.38	51.8
1	16.00	0	0	0.6679	0	3	0	3	500	PLASMA	73.2	173	31	2	7.32	51.8
1	24.00	0	0	0.3158	0	3	0	3	500	PLASMA	73.2	173	31	2	7.22	51.8
2	0.00	500	0	0.0000	1	1	1	3	500	PO_DOSE	63.1	168	41	2	6.96	46.6
2	0.50	0	0	7.6662	0	3	0	3	500	PLASMA	63.1	168	41	2	6.96	46.6
2	1.00	0	0	7.0924	0	3	0	3	500	PLASMA	63.1	168	41	2	6.96	46.6
2	2.00	0	0	6.4259	0	3	0	3	500	PLASMA	63.1	168	41	2	6.95	46.6
2	4.00	0	0	4.3544	0	3	0	3	500	PLASMA	63.1	168	41	2	6.94	46.6
2	6.00	0	0	2.9303	0	3	0	3	500	PLASMA	63.1	168	41	2	6.93	46.6
2	8.00	0	0	2.7519	0	3	0	3	500	PLASMA	63.1	168	41	2	6.92	46.6
2	12.00	0	0	1.5452	0	3	0	3	500	PLASMA	63.1	168	41	2	6.90	46.6
2	16.00	0	0	0.8061	0	3	0	3	500	PLASMA	63.1	168	41	2	6.88	46.6
2	24.00	0	0	0.3172	0	3	0	3	500	PLASMA	63.1	168	41	2	6.84	46.6
3	0.00	500	0	0.0000	1	1	1	3	500	PO_DOSE	55.5	160	41	2	7.14	41.6
3	0.50	0	0	1.0684	0	3	0	3	500	PLASMA	55.5	160	41	2	7.13	41.6
3	1.00	0	0	2.0173	0	3	0	3	500	PLASMA	55.5	160	41	2	7.11	41.6
3	2.00	0	0	9.2194	0	3	0	3	500	PLASMA	55.5	160	41	2	7.08	41.6
3	4.00	0	0	5.0079	0	3	0	3	500	PLASMA	55.5	160	41	2	7.02	41.6
3	6.00	0	0	3.7046	0	3	0	3	500	PLASMA	55.5	160	41	2	6.96	41.6
3	8.00	0	0	2.7111	0	3	0	3	500	PLASMA	55.5	160	41	2	6.90	41.6
3	12.00	0	0	1.4441	0	3	0	3	500	PLASMA	55.5	160	41	2	6.79	41.6
3	16.00	0	0	0.7142	0	3	0	3	500	PLASMA	55.5	160	41	2	6.68	41.6
3	24.00	0	0	0.2693	0	3	0	3	500	PLASMA	55.5	160	41	2	6.47	41.6
4	0.00	500	0	0.0000	1	1	1	3	500	PO_DOSE	67.0	167	30	2	8.52	47.9
4	0.50	0	0	0.3319	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	1.00	0	0	2.2650	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	2.00	0	0	6.0122	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	4.00	0	0	3.8868	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	6.00	0	0	2.8472	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	8.00	0	0	2.2171	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	12.00	0	0	1.4061	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	16.00	0	0	0.6193	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
4	24.00	0	0	0.2279	0	3	0	3	500	PLASMA	67.0	167	30	2	8.52	47.9
5	0.00	500	0	0.0000	1	1	1	3	500	PO_DOSE	52.5	165	29	2	5.77	41.2
5	0.53	0	0	2.0682	0	3	0	3	500	PLASMA	52.5	165	29	2	5.78	41.2



## Résumé

Sanofi-Aventis est le premier groupe pharmaceutique en Europe et le troisième sur l'échelle mondiale. La recherche est l'axe stratégique que développe cette entreprise afin de participer à l'amélioration de la santé humaine. J'ai réalisé mon stage de 1<sup>re</sup> année de Master dans ce secteur et plus précisément dans le service de Pharmacocinétique/Pharmacodynamie des populations basé à Francfort (Allemagne).

Ce pôle a pour objet l'étude de la variabilité de réponse des individus à une dose donnée dans le but d'améliorer la posologie d'un médicament. Ceci implique l'utilisation de modèles et de programmes informatiques de régression mathématique tels que NONMEM.

Après l'apprentissage des différentes méthodes et outils utilisés par l'équipe, j'ai participé en partenariat avec un autre stagiaire de l'I.U.P. à l'optimisation de ces méthodes et à la création de deux nouvelles. Variation.Onetheta permet l'étude à partir de la variation d'un seul des paramètres initiaux. Simplefit extrait chaque individu d'un échantillon et lance une analyse sur chacun d'eux. Ces deux méthodes rassemblent les données afin de présenter les résultats sous forme de rapport.

## Abstract

Sanofi-Aventis is the first pharmaceutical group in Europe and the third worldwide. Research is the strategic axis which develops this company in order to participate in the improvement of human health. I carried out my training course of first year of Master in this sector and more precisely in the service of population Pharmacokinetic/Pharmacodynamic modeling located in Frankfurt (Germany).

This pole aims to, the study the variability of answer of the individuals to an amount given in order to improve posology of a drug. This implies the use of models and programs of mathematical regression such as NONMEM.

After the training of the various methods and tools used by the team, I took part in partnership with another trainee of the I.U.P. with the optimization of these methods and creation of two news. Variation.Onetheta allows the study starting from the variation of only one of the initial parameters. Simplefit extracts each individual from a sample and launches an analysis on each one of them. These two methods gather the data in order to present the results in the form of report.